Making Data Work for You: Data Warehousing in the Cloud

Colin White
BI Research
February 24, 2015
Speakers

Colin White
President,
BI Research

Andy Ellicott
VP, Product Marketing,
IBM Cloudant
Making Data Work For You: Data Warehousing in the Cloud

Colin White
President, BI Research
TDWI and IBM Webinar
February 2015
Introduction

Cloud computing is evolving rapidly and vendors are delivering a range of products and services for enabling DW in the cloud.

Given the many cloud offerings on the market and it is essential that you match requirements to the appropriate offering.

Not all DW projects are suited to a cloud approach, and it is important to identify those projects that do and do not lend themselves to the cloud environment.

The objectives of this webinar are to provide an overview of industry trends in the use of DW in the cloud and to identify the use cases that best fit a cloud environment.
Building an Enterprise DW Today is Difficult

Multiple data sources

- Operational data
- DW historical data
- Text & media files
- Web & social content
- Sensor data

Increasing data volumes & data rates

Multiple analytic tools

- Data integration
- Data management
- Data analysis
- Decision management

Multiple deployment options

Multiple output formats

Multiple user devices
Why? The Traditional DW Architecture

Data must be modeled, acquired, integrated and loaded into a DW before it can be analyzed.
Solution: Modernize the DW Architecture

Traditional EDW environment

Investigative computing platform

Data integration platform

Data refinery

Other internal & external structured & multi-structured data
Real-time streaming data

RT BI services
Operational systems

RT analysis platform
Operational real-time environment

Analytic tools & applications
A New Generation of Data Warehousing

**DRIVERS**
- New business insights
- Reduced costs

**TECHNOLOGIES**
- New technologies
- Enhanced data management
- Enhanced analytics
- New deployment options

Next generation DW
New Technologies

Enhanced Data Management
- New data sources (big data)
- Analytic relational DBMSs
- Non-relational systems (e.g., Hadoop, HBase, MongoDB, CouchDB)
- Improved price/performance

Enhanced Analytics
- Investigative computing
- Predictive & prescriptive analyses
- Enhanced visualization

New Deployment Options
- Integrated H/W & S/W appliances
- Cloud computing
- Mobile devices (mobile first)
IBM Perspective on the Cloud

“Data is the WHAT
Cloud is the HOW
Insight is the WHY”

Bob Picciano, Senior VP of the Information & Analytics Group, IBM Insight 2014
The Role of Cloud Computing

Provides pay-as-you go, on demand and elastic scalability for prototyping, developing and deploying many IT projects.

Reduces upfront IT costs and enables companies to scale IT resources as required, while paying only for the resources used.

Cloud technologies are maturing and organizations are overcoming their data security issues and concerns.

Key issue is the complexity of integrating cloud and on-premises data and the inability of many cloud services to efficiently and rapidly move data into and out of the cloud environment.
Most traditional DW projects to date have involved managing and analyzing data extracted from on-premises operational systems. Private clouds are being used to consolidate data, applications, and servers to reduce costs and improve governance. Public cloud services are being used for delivering analytics on data stored in the cloud such as operational data from Salesforce.com.

- Usually standalone and piecemeal projects
- Exception is in Web-focused companies whose main business and data are oriented towards Internet commerce

One of the risks of public cloud computing is that it makes it easier for business groups to bypass IT and purchase their own solutions. Important for IT to partner with the business in deploying cloud services to reduce risk, avoid poor technology selection, and manage data governance and data security requirements.
DW in the Cloud: The Potential

The biggest potential for cloud computing is the processing of data that already exists in the cloud.

Includes the large volumes of data on internal and public web servers, externally generated data (e.g., from sensors) and also data generated by third-party providers.

These large volumes of data can be captured and transformed in the cloud and then delivered to an in-house system for analysis.

In many cases, the data can also be analyzed in the cloud and the results delivered to internal and external business users.

One of the key requirements here is to keep data movement to a minimum and to process data where it resides.

Look for projects and solutions that not only simplify development, deployment and administration, but that also provide solid and well performing data integration and data movement capabilities.
Deploying in Cloud: Potential Use Cases

1. Traditional EDW environment
2. Investigative computing platform
3. Data integration platform
4. Data refinery
5. Other internal & external structured & multi-structured data
6. Real-time streaming data
7. Operational systems
8. RT analysis platform
9. RT BI services
10. Operational real-time environment
Selecting Cloud Computing Services

When cloud services initially became available, the market was considered to consist of three main types of service:

- **Software-as-a-service (SaaS)**
- **Platform-as-a-service (PaaS)** – may also see subsets such database-as-a-service (DBaaS)
- **Infrastructure-as-a-service (IaaS)**

As the use of these services has grown and more vendors have entered the market, this classification scheme is often inadequate.

A variety of new schemes have emerged, but many of these confuse customers and make vendor selection overly complex.
Forrester Research Perspective

Focus should be on the developer (and uses cases and users)

Source: The Forrester Wave: Enterprise Public Platforms, Q2 2013
End-To-End Solution: Analytics as a Service

Need to consider requirements across the complete information supply chain — from data acquisition to data delivery.

By developing a comprehensive AaaS cloud-based data strategy, companies can define an integrated framework and optimize the value of data.

An AaaS insight framework (IaaS+PaaS+SaaS) encompasses:

- Capturing and extracting structured and unstructured data from trusted sources, including prioritizing the most critical data and identifying what to retain and for how long.
- Managing and controlling data under policy and governance guidelines across the enterprise and in compliance with industry requirements.
- Performing data integration, analysis, transformation, and visualization to deliver the right information to the right location at the right time.

*Source: “Big Data in the Cloud: Converging Technologies.” Intel Solution Brief, February 2013*
Aid the implementation, administration and support of an end-to-end solution that covers the complete data lifecycle from data acquisition to data analysis and delivery of results to the user.

Some cloud vendors provide managed services as part of their offerings – the issue here is that the term managed services is used differently by vendors.

In many cases these services are often technology or platform related and are not always ideally suited to DW projects.

During vendor evaluations, it’s important to understand exactly what a vendor means by the term “managed service”.

- **Monitoring** – preserving the security and availability of your cloud infrastructure
- **Break-fix** – incident lifecycle management and root cause analysis for continuous improvement
- **Maintenance** – ensuring your cloud infrastructure, operating systems, middleware and applications are kept up-to-date
- **Reporting** – periodic reporting of cloud infrastructure performance against contractual Service Level Agreements
- **Optimization** – regular review of cloud service usage and capacity to aid infrastructure right-sizing
- **Provisioning** – spin-up or tear-down of infrastructure on behalf of the customer
Getting Started: Standalone Solution

Standalone reporting and analysis of Web, social media and/or sensor data: Data from each source is uploaded in small batches or streamed directly to the cloud service for reporting and analysis.
Data analysis and visualization of e-commerce data: A cloud-based system is ideally suited to collecting, managing, analyzing and visualizing data that is often deployed on hundreds of web servers that handle requests from millions of users on a variety of devices.
Data warehouse augmentation: A cloud-based data refinery is a cost-effective way of capturing, storing, transforming and archiving data while also providing connectivity to a DW.

The on-premises EDW remains the primary source of analytics for users.

The analysis of cloud-based data may also be provided as required for investigative purposes.
Thanks for Listening

GET ALL THE INFORMATION YOU CAN, WE'LL THINK OF A USE FOR IT LATER.
IBM Cloud Data Warehousing Services

Andy Ellicott, VP Product Marketing, IBM Cloudant
IBM dashDB – Data Warehousing as a Service

Terabytes of data ready to analyze within minutes

• Fast querying
  • In-memory
  • Columnar
  • SIMD hardware acceleration
  • Actionable compression

• Advanced analytics
  • Support for OLAP SQL extensions
  • In-database analytics & R for predictive modeling

• Easy to integrate
  • Connect common 3rd party BI tools

keeps data warehouse infrastructure out of your way
IBM BigInsights Cloud – Enterprise Hadoop as a Service

Big Data – easier, faster, richer

- Analyze petabytes of any kind of data on elastic compute clusters

- Easy to adopt
  - Provisioned, hosted, scaled for you

- Native Hadoop enriched with:
  - Faster performance
    - ~4x faster than open source Hadoop
  - Visualization tools
  - Built-in R, streaming & text analytics
  - Data protection & governance

- Easy to integrate
  - Connect common 3rd party BI tools

IBM BigInsights Cloud

Open source Hadoop components

Enterprise Capabilities

Visualization & Exploration
Development Tools
Advanced Engines
Connectors
Workload Optimization
Administration & Security
IBM Cloudant NoSQL DBaaS

- JSON document database

- For apps that need:
  - Elastic scalability
  - High availability
  - Data model flexibility
  - Data mobility
  - Text search
  - Geospatial

- Not normally a DW, but good landing & transformation area for JSON
  - Updates views incrementally, in seconds-minutes, versus hours in RDBMS

No-Brainer Cloud Data Warehousing Use Cases

- **Agile analytics**
  - Analysts create data marts on a whim
  - For short-lived or spontaneous DW projects (e.g., re-pricing)
  - Need it today? It’s ready today…with unlimited scalability
  - Kill it (and cost) when you’re done
  - Every analyst should have this ability

- **Analytic SaaS apps**
  - Cloud apps that embed analytics
  - Put your DW in the cloud, close to the app code that accesses it
  - Better performance, lower bandwidth costs
  - Easier access by customers, partners,…
Somewhat-Brainer Cloud Data Warehousing Use Case

The “lower TCO” use case

- **Cloud vs. on-premise TCO comparison:**
  - Monthly DWaaS operational expense x life expectancy (months)
  - Up-front capital investment in new DBMS hardware, software, annual maintenance, DBA (redesign costs as it scales)

- **Your mileage will vary**
  - Depending on on-premise cost & life expectancy (monthly fee) of cloud DW
5 Cloud Data Management Red Herrings

- **Security?**
  - Security of on-premise DBMS (DB2) is implemented, plus at-rest & in-flight encryption usually

- **Integration?**
  - Standard DB connectivity (ODBC, JDBC, et al) is supported & ETL also tends to support web targets

- **Performance?**
  - State of the art analytic DB architecture, running on bare metal hardware*.

- **Cloud lock-in?**
  - Not proprietary data formats; easy enough to migrate data to on-premise (or other cloud) RDBMS
Try It for Free…

http://bluemix.net
Thank You
Questions?
Contact Information

If you have further questions or comments:

Colin White, BI Research
info@bi-research.com

Andy Ellicott, IBM
aellicot@us.ibm.com